A Proposed High-Frequency, Negative-Resistance Diode

By W. T. READ, JR.

(Manuscript Received March 2, 1957)

This paper describes and analyzes a proposed semiconductor diode designed to operate as an oscillator when mounted in a suitable microwave cavity. The frequency would be in the range extending from 1 to 50 kmc. The negative Q may be as low as 10 and the efficiency as high as 30 per cent.

The diode is biased in reverse so as to establish a depletion, or space-charge, layer of fixed width in a relatively high resistance region, bounded by very low resistance end regions. The electric field has a maximum at one edge of the space-charge region, where hole-electron pairs are generated by internal secondary emission, or avalanche. The holes (or electrons) travel across the space-charge layer with constant velocity, thus producing a current through the diode. Because of the build-up time of the avalanche, and the transit time of the holes across the depletion layer, the alternating current is delayed by approximately one-half cycle relative to the ac voltage. Thus, power is delivered to the ac signal. When the diode is mounted in an inductive microwave cavity tuned to the capacity of the diode, an oscillation will build up. It appears possible to obtain over 20 watts of ac power in continuous operation at 5 kmc.

I. DESCRIPTION

This paper discusses a proposed oscillator consisting of a semiconductor diode biased in reverse and mounted in a microwave cavity. The impedance of the cavity is mainly inductive and is matched to the mainly capacitative impedance of the diode so as to form a resonant system. We shall show that the diode can have a negative ac resistance so that it delivers power from the dc bias to the oscillation. The negative Q may be as low as 10 and the efficiency as high as 30 per cent.

The principle of operation is as follows: a reverse bias is applied to establish a space-charge, or depletion, layer of fixed width in a relatively weakly doped region bounded by highly doped end regious. A possible

structure is the n^+-p-i-p^+ structure, where the + denotes high doping and i means intrinsic. This structure is shown in Fig. 1(a). The field distribution under reverse bias is shown in Fig. 1(b). The voltage is always well above the punch-through voltage, so that the space-charge region always extends from the n^+-p junction through the p and i (intrinsic) regions to the i-p^+ junction. The fixed charges in the various regions are shown in Fig. 1(b). A positive charge gives a rising field in going from left to right. A positive field makes holes move to the right. The maximum field, which occurs at the n^+-p junction, is of the order of several hundred kilovolts per cm, so that hole-electron pairs are generated by internal secondary emission (also called multiplication or avalanche). The electrons go immediately into the n^+ region. The holes move to the right across the space-charge region. The field throughout the space-charge region is in the range (above about 5 kilovolts per cm) where carriers move with constant velocity independent of the field. For practical purposes we can forget about the electrons in the space-charge

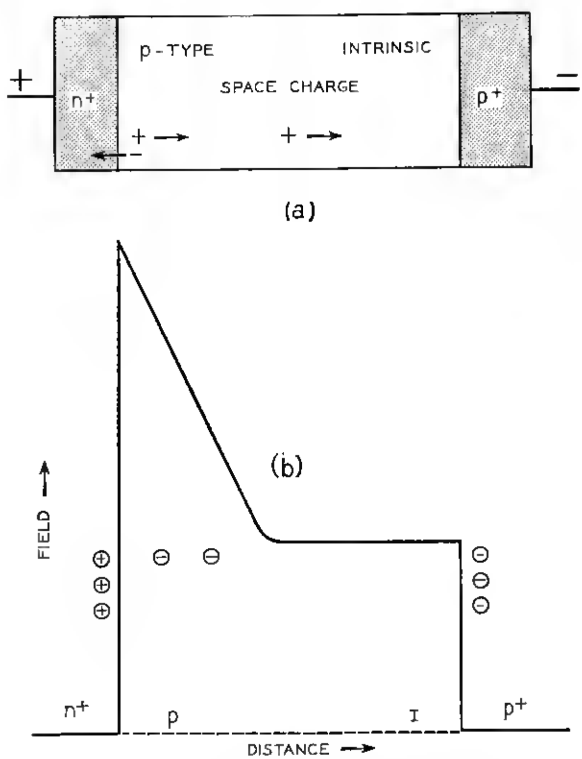

Fig. 1 — The structure (a) and field distribution (b) under reverse bias.

region and consider that current is generated at the n^+-p junction, so that only the holes move in the space-charge region. Thus, the physical picture is as follows: a current of holes $I_0(t)$ is generated at the n^+-p junction. The holes move to the right and traverse the space-charge region, moving with constant velocity, v. In silicon v is about 10^7 cm sec^{-1}. The transit time of a hole across the space-charge region is $\tau = W/v$, where W is the width of the space-charge region. Throughout the discussion we shall illustrate various quantities by giving numerical values for the case $W = 10^{-3}$ cm. Thus τ would be 10^{-10} sec.

The holes moving across the space-charge region produce a current $I_e(t)$ in the external circuit. It can be shown that I_e is equal to the average current in the space-charge region. Since velocity is constant, I_e is simply $v/W = 1/\tau$ times the total charge of the moving holes. (We shall assume throughout that all quantities refer to unit area of junction.) Suppose, for example, that a pulse of holes of charge δQ is suddenly generated at the n^+-p junction. Immediately, a constant current $I_e = \delta Q/\tau$ begins to flow in the external circuit, and continues to flow during the time, τ, that the holes are moving across the space-charge region from the n^+-p junction to the p^+ region. Thus, on the average, the external current $I_e(t)$ due to the moving holes is delayed by $\tau/2$ relative to the current $I_0(t)$ generated at the n^+-p junction. We shall show (in the discussion of multiplication) that the current $I_0(t)$ is delayed by one-quarter of a cycle relative to the ac voltage. Thus, to get a total delay of one-half cycle, we want the delay $\tau/2$ due to transit time to be one-quarter of a cycle. The cavity should therefore be tuned to give a resonant frequency $(2\tau)^{-1}$.

The conductive current $I_e(t)$, which arises from carriers moving through the space-charge layer, should be distinguished from the displacement, or capacitative, current I_c, which charges and discharges the diode regarded as a capacitor. This current, I_c, supplies the variation in charge at the edges of the space-charge region, where the field changes abruptly and, for practical purposes, can be considered discontinuous. Since I_c is 90° out of phase with the voltage it contributes nothing to the power.

1.1 *Multiplication*

Carriers moving in the high field near the n^+-p junction acquire enough energy to knock valence electrons into the conduction band, thus producing hole-electron pairs. The rate of pair production, or multiplication, is a sensitive nonlinear function of the field. By proper doping, the field can be given a relatively sharp peak so that multiplication is confined to a very narrow region at the n^+-p junction. The multiplication

rate can be regarded as a function of the peak field, E_0. At a critical field, E_c, breakdown will occur; that is, any current will be self-sustaining; every pair produced will on the average produce one other pair. We shall neglect the thermally generated reverse saturation current, which will be small compared to the generated current $I_0(t)$. Then, when the field is above E_c, the current $I_0(t)$ will be more than self-sustaining and will build up. When the field is below the critical breakdown field E_c, the current is less than self-sustaining and will die down.

In operation, the diode is biased so that the peak field is above E_c during the positive half of the voltage cycle and below E_c during the negative half. Hence the current $I(t)$ builds up during the positive half and dies down during the negative half. Therefore $I_0(t)$ reaches its maximum in the middle of the voltage cycle, or one-quarter of a cycle later than the voltage.

We have assumed so far that the field varies in phase with the voltage. This will be a good approximation if the carrier space-charge can be neglected. This can be seen from Fig. 1(b). If the voltage is above the punch-through voltage, an increase in voltage simply raises the whole field distribution throughout the depletion layer. That is, additional charges simply appear at the edges of the space-charge layer. Hence the field at each point varies in phase with the voltage. The space charge of the carriers will, however, affect the shape of field distribution in the space-charge region. When the current becomes too large, the carrier space-charge cannot be neglected. As we shall see, the effect of carrier space-charge is to reduce the delay between the voltage and the current generated, I_0. In practice this limits the dc bias current.

When the dc bias is small enough so that carrier space charge can be neglected, the operation can be summarized as follows: the peak field varies in phase with the voltage and generates (at one edge of the depletion layer) a current delayed by 90°. This current gives rise to a current through the diode delayed by 90° relative to the current generated and, therefore, by 180° relative to the voltage.

1.2 *Diode and Cavity*

The capacity of the diode and the inductance of the cavity determine the frequency, f, of oscillation. We have seen that the optimum frequency, f, is such that the transit time, τ, is one-half cycle. Hence the angular frequency, ω, should be

$$\omega = 2\pi f = \frac{\pi}{\tau} = \frac{\pi v}{W}. \tag{1}$$

For $W = 10^{-3}$ cm, the frequency, f, would be 5,000 mc.

We have seen that the current into the diode is $I_e + I_c$, where I_e is a conductive current due to holes moving in the space-charge region and I_c is a displacement current required to charge the diode regarded as a capacitor. Thus, the diode acts like a capacity and negative resistance in parallel.

The capacity is related to the width, W, of the depletion layer by

$$\begin{aligned} C &= \frac{\kappa}{4\pi W}, \\ &= \frac{1.06 \times 10^{-12}}{W} \quad \text{for Si,} \end{aligned} \tag{2}$$

where C is in farads per cm^2 and W in cm. The negative conductance of the diode will be small compared to ωC, so the admittance is mainly that of the capacity. Thus for $W = 10^{-3}$ cm the admittance will be about 30 mhos per cm^2.

To make a resonant system we mount the diode in a microwave cavity having a mainly inductive impedance. The inductance, L, of the cavity is chosen so that the cavity and diode together have a resonant frequency $f = \frac{1}{2}\tau$.

Fig. 2 shows a possible cavity with the diode mounted in it. The diode should have narrow end regions so that the depletion layer is as close as possible to the metal base and center post of the cavity. Then the heat generated will flow away rapidly. The cavity is tuned by adjusting the vertical position of the plunger forming the top of the cavity. There must be some break in the cavity so that a dc bias can be applied. This is accomplished in Fig. 2 by insulating the center lead. The break could also be made by insulating the walls from the base. On the right is an outlet through which the ac power is tapped.

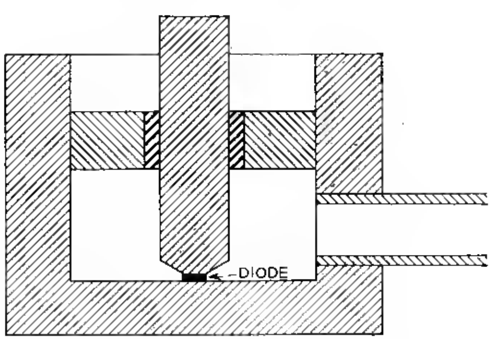

Fig. 2 — Diode mounted in microwave cavity.

The impedance, ωL, of the cavity must be equal to the impedance of the diode. If the diode is circular with radius R, then the junction area is πR^2 and the impedance, $(\pi R^2 \omega C)^{-1}$. Thus the area of the diode is limited by how small the impedance of the cavity can be made. Setting $\omega L = (\pi R^2 \omega C)^{-1}$ and using (1) and (2) we have

$$\begin{aligned} \left(\frac{R}{W}\right)^2 &= \frac{4}{\pi \kappa v \omega L}, \\ &= \frac{9.5 \times 10^3}{\omega L} \quad \text{for Si,} \end{aligned} \tag{3}$$

where ωL is in ohms. It would be difficult to make ωL less than about 10 ohms. Hence R could not be more than about $30W$.

1.3 *Stable Operation*

We have shown that the diode will deliver power, so an oscillation will build up. In Section III we shall calculate exactly the Q for the small signal case where everything is linear. For reasonable values of the direct current bias, I_d, the Q can be made negative and as low in magnitude as desired. As the oscillation builds up, the behavior becomes highly nonlinear. In the practical range of operation the field will vary by a fraction — probably less than 20 per cent — of its average value but the current will vary by orders of magnitude. The current I_e approaches a square wave, being negligibly small during the positive half of the ac voltage cycle and almost constant during the negative half cycle. Since the direct current I_d is the average conductive current, it follows that the amplitude of variation of I_e is approximately equal to I_d. If V_a is the ac voltage amplitude, the ac power delivered P_r is found to be

$$P_r = \frac{2}{\pi} I_d V_a \tag{4}$$

per unit area of junction. Thus, if the dc bias is applied by a constant direct current generator, the power delivered is proportional to V_a. The energy stored in the capacity is proportional to V_a^2. Hence, $-Q$ is proportional to V_a. In other words, if the amplitude increases, the stored energy, or energy of oscillation, increases faster than the energy delivered per cycle. This is the condition required for a stable oscillation to be possible. At the stable operating point, $-Q$ of the diode is equal to the effective Q of the cavity. If the amplitude increases, the energy delivered by the diode increases less than the energy lost to the cavity. Thus, the amplitude decreases. In the same way a decrease in amplitude is also self-correcting.

The total power output is P_r times the area πR^2 and is therefore pro-

portional to $R^2 I_d V_a$. The maximum R is given by (3). The maximum allowable I_d is limited by the fact that carrier space-charge cannot be too large, as we now show.

1.4 *Space Charge of the Carriers*

The holes moving across the space-charge region will affect the field distribution. As shown in Fig. 1(b) a positive charge gives a rise in field (in going from left to right) and a negative charge, a decrease in field. Thus, the effect of the holes will be to oppose the fixed negative charge in the depletion layer, and hence to flatten the field distribution shown in Fig. 1(b). Therefore, for a given voltage, the effect of the holes is to reduce the peak field at the n^+-p junction. The multiplication rate increases with the peak field. Thus, the current generated has a space charge that tends to reduce the rate of current generation. That is, the current tends to shut itself off. Instead of building up throughout the positive half of the voltage cycle, $I_0(t)$ builds up until the carrier space-charge has reduced the peak field below the sustaining field E_c. Then the current decreases. Thus, I_0 reaches its peak *before* the middle of the cycle. This reduces the delay and hence the power. Increasing the current, therefore, increases the ac power only up to a certain point where carrier space-charge begins to spoil the phase relations.

We have seen that in the practical range of operation the current I_e varies by approximately its dc value I_d. A current I_d corresponds to a total carrier charge τI_d in the space-charge layer (since I_e is the average current in the space-charge layer and all carriers travel with velocity $v = W/\tau$). If the carrier space-charge is to be neglected it must be small compared to the charge CV_a that produces the voltage variation. Thus we want $\tau I_d \ll CV_a$. Actually, as we shall show in Section III, it is sufficient to take $I_d \tau = CV_a/2$. Since the period of oscillation is 2τ, this becomes

$$I_d = \frac{\omega}{2\pi} CV_a. \tag{5}$$

If I_d is no greater than this the current and voltage will be roughly 180° out of phase and we can use (4) for the power per unit area. For larger I_d the increase in power would be more than offset by the effect of carrier space-charge on the phase relations.

1.5 *Power Output*

Combining (3) with (5) we have, for the total ac power output

$$\pi R^2 P_r = \frac{V_a^2}{\pi^2 \omega L}. \tag{6}$$

As we have seen, the impedance ωL of the cavity could be as low as 10 ohms. Now consider the maximum V_a. This is limited by the dc voltage V_d. It is seen from Fig. 1(b) that if V_a is too large the field in the intrinsic region will be reduced to zero during the negative voltage cycle. By careful doping, the voltage at punch-through can be made small compared to the dc bias at breakdown. Even allowing for the reduction in field due to the carrier space-charge, we will be safe with $V_a = V_d/2$. This gives an efficiency of about 30 per cent. From (6) the power becomes

$$\text{Power Output} = \pi R^2 P_r = \left(\frac{V_d}{2\pi}\right)^2 \frac{1}{\omega L}. \tag{7}$$

The dc voltage is limited by the following considerations: (a) In the negative half cycle the field must not fall below about 5 kilovolts per cm. Otherwise carrier velocity will depend on field; so the carriers will slow down during the negative half of the ac voltage cycle and thus reduce the power. This can be avoided if V_d is at least $10^4 W$ volts, where W is in cm. (b) To localize the multiplication the field must remain well below E_c except near the n^+-p junction. This will be so if $V_d = 2V_a$ is less than about $(2/5)E_c W$. Then the maximum field in the intrinsic region will always be less than 0.6 E_c. Thus we have $10^4 W < V_d < 0.4\, E_c W$.

For W as large as 10^{-3} cm the critical field will be about 350 kilovolts per cm. For W as low as 10^{-4} cm the multiplication would have to be confined to a region no more than 10^{-5} cm in width. This would require a somewhat steeper field gradient and a higher critical field — say about 650 kilovolts per cm. Thus for $W = 10^{-4}$ cm, or 50 kmc, V_d would have to be less than 26 volts. For $\omega L = 10$ ohms, the maximum power output would then be less than 2 watts. At $W = 10^{-3}$ cm, or 5 kmc, the maximum V_d would be 140 volts, which gives a power of 50 watts. At $W = 10^{-2}$, or 500 mc, the maximum voltage would be 1400 volts, and the minimum 100 volts. The power would then be between 25 watts and 5 kw.

1.6 *Zener Current*

Chynoweth and McKay[1] have shown that in sufficiently narrow junctions, where breakdown occurs at around 10 volts or less, the current is generated not by multiplication but by internal field emission. That is, electrons tunnel from the valence to conduction bands. This is known as Zener current. We shall show in Section IV and Appendix E that the device will operate on Zener current, but much less effectively. This,

even more than the limited power, may limit the minimum W and hence the maximum frequency that would be practical.

Experiments

An experimental program has been undertaken to construct and test a device operating by avalanche.

II. CURRENT AND SPACE CHARGE

In this section we consider the physics of multiplication, space-charge and carrier flow in more detail and obtain the equations that govern the behavior of the diode.

2.1 *Multiplication*

Electrons moving with velocity v generate hole-electron pairs at a rate of $\alpha v n$, where n is the electron concentration and α is the ionization rate. By definition α is the number of pairs produced on the average by an electron in moving unit distance. Thus α^{-1} is the average distance between ionizations. We shall be dealing with the case where each pair produces roughly one other pair; thus α^{-1} is of the order of the width of the narrow region near the n^+-p junction where the multiplication is localized. For fields below about 600 kilovolts per cm α can be considered a function of field, $\alpha = \alpha(E)$. For larger fields the electrons do not have time to get into equilibrium with the field in the average distance α^{-1} between collisions. We shall be dealing with fields where $\alpha = \alpha(E)$.

McKay[2] has measured α as a function of E for electrons in silicon. Fig. 3 is a plot of the best fit curve to McKay's data, together with the theoretical curve calculated by P. Wolff[3]. A straight line of slope 6 is also shown and is seen to be a good fit to the data below about 500 kilovolts per cm. In the formulas we shall use the general relation $\alpha \approx E^m$ and take $m = 6$ for numerical calculations.

Assumptions

We shall make the simplifying assumption that $\alpha = \alpha(E)$ is the same for both holes and electrons. Actually this is not so. However, α is so sensitive to field that the difference in field for the same α is small. Any difference in the α's will alter both the carrier and field distributions so as to favor the carrier with the lower α. Consequently, it is believed that this assumption will not give misleading results.

We also assume that holes and electrons travel with the same velocity in the high fields involved. This will not lead to serious error since the motion of one type of carrier always plays a minor role.

2.2 *DC Case*

We take the x axis normal to the junctions with $x = 0$ at the n^+-p junction and $x = W$ at the i-p^+ junction, as shown in Fig. 4. It can be shown that the direct current I is related to the thermally generated re-

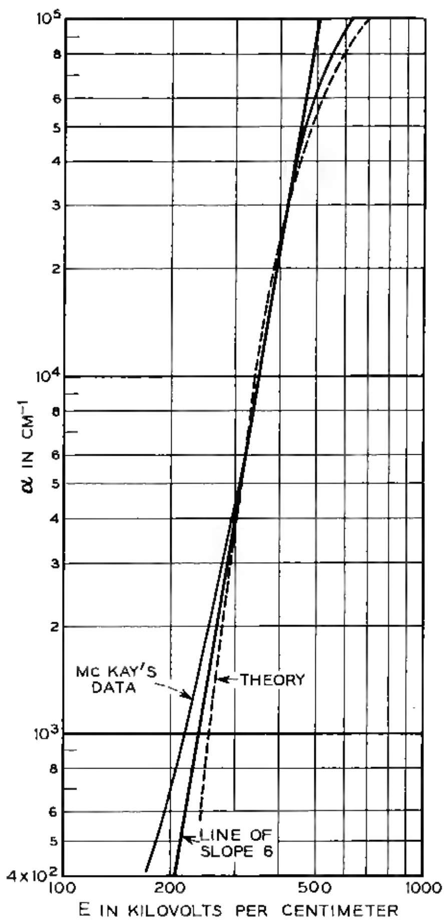

Fig. 3 — Ionization rate vs electric field.

verse saturation current I_s by

$$\frac{I_s}{I} = 1 - \int \alpha(E)\ dx \tag{8}$$

where the integral is taken over the space-charge region. See, for example, McKay[2]. We will be dealing with space-charge regions where most of the multiplication occurs in a narrow region, which we shall call the *multiplication region*, near $x = 0$. Thus we need take the integral in (8) only from $x = 0$ to $x = x_1$, where $x_1 \ll W$ is the width of the narrow multiplication region.

Breakdown occurs, that is, the direct current becomes infinite, when $\int \alpha\, dx = 1$. Physically this means that each hole-electron pair generated in the multiplication regions will, on the average, generate one other pair. We now consider how $\int \alpha\, dx$ depends on the peak field E_0 for various field distributions including the one shown in Fig. 4.

(a) The simplest case is that where the field is a constant E_0 in the narrow multiplication region. This would correspond to a short region of high constant field followed by an abrupt drop to a much lower field.

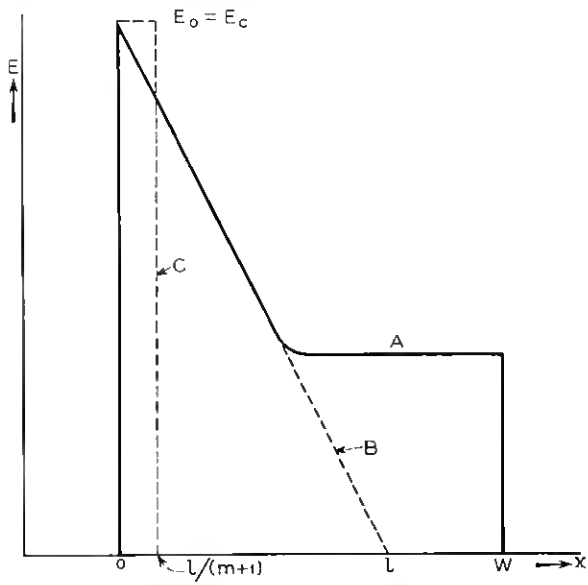

Fig. 4 — Field distribution at breakdown.

Then

$$\int \alpha \, dx = \int_0^{x_1} \alpha(E_0) \, dx = x_1\alpha(E_0).$$

So if α is proportional to E^m

$$\int \alpha \, dx = \left(\frac{E_0}{E_c}\right)^m,$$

where E_c the critical field for breakdown is given by $\alpha(E_0) = 1/x_1$.

(b) Next consider the linear field distribution shown in Fig. 4. Curve A is the field distribution at breakdown. We can take $E = E_0 - kx$ in the multiplication region. The slope k is proportional to the charge in the p-region. We shall neglect the effect of the carriers on the space charge. Then k will be proportional to the fixed charge in the p-region and will be constant independent of current. In practice k would be well above E_c/W. In order for the carriers to produce a field gradient comparable to this, the current would have to be of the order of $(\kappa v/4\pi)(E_c/W)$, or several thousand amperes/cm^2. Actual currents will be much smaller than this.

In Fig. 4 the flat section of field will contribute negligibly to the multiplication, so we can replace the actual field distribution A by the tangent curve B. Then

$$\int \alpha \, dx = \left(\frac{E_0}{E_c}\right)^{m+1}, \tag{9}$$

where E_c is given by $\alpha(E_c) = (m + 1)/l$ and $l = E_c/k$ is the zero intercept of the tangent curve B at breakdown. Thus the peak field E_c at breakdown is equal to the breakdown field for a constant field region of width $l/(m + 1)$, as shown by curve C in Fig. 4.

(c) For a linear gradient junction the field varies parabolically with x and

$$\int \alpha \, dx = \left(\frac{E_0}{E_c}\right)^{m+1/2}.$$

In what follows we shall use (9), which applies to a linear field, or abrupt junction. This will be perfectly general if m is left arbitrary. The results obtained for $m = 6$ will be a good approximation for a linear gradient junction, since the difference between 7 and 6.5 is within the experimental uncertainty in m.

As an example, consider a step junction with $m = 6$ and $W = 10^{-3}$ and let $l = 0.7W$. Then the critical field E_c is given by $\alpha(E_c) = 7/l = 10^{-4}$ cm. From Fig. 3, $E_c = 350$ kilovolts/cm.

2.3 *Time Dependence*

In the dc case $\int \alpha\, dx$ cannot be greater than unity. That is, we cannot get above breakdown. This is not necessarily so for rapidly varying fields. We now derive a differential equation for the current as a function of time. This will reduce to (8) in the dc case.

Let p and n be the hole and electron densities respectively. The corresponding currents are $I_p = qvp$ and $I_n = -qvn$, where q is the electronic charge. Hole-electron pairs are being generated at a rate $\alpha v(n + p)$. This is so large compared to the rate of thermal generation that the latter can be neglected. The equations of continuity then become

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\frac{\partial I_p}{\partial x} + \alpha v(n + p), \tag{10}$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial I_n}{\partial x} + \alpha v(n + p), \tag{11}$$

where $\alpha = \alpha(E)$. The three variables, p, n and E are determined by (10) (11) and by Poisson's equation and the boundary conditions at the junctions. An exact solution is impractical. Instead, we make an approximation based on the fact that most of the generation is confined to the narrow multiplication region near $x = 0$. This is shown in Fig. 5. The multiplication region extends from $x = 0$ to $x = x_1$. The width x_1 is a small fraction of the width W of the space-charge region.

2.4 *Assumptions*

We shall assume that (a) *all* of the multiplication occurs in the relatively narrow multiplication region and (b) the total current $I(x, t) = I_p(x, t) + I_n(x, t)$ in the multiplication region is a function $I(x, t) = I_0(t)$

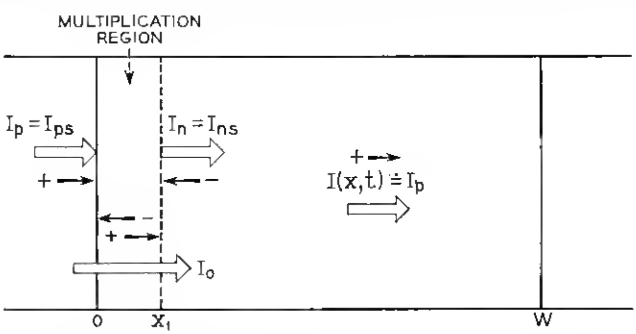

Fig. 5 — Boundary conditions on the multiplication region.

of time only. This, of course, will not be true of I_p and I_n individually in either the ac or dc cases. Assumption (a) will be valid if the net acceptor concentration in the p-region is sufficient to give a sharp drop in field and hence to localize the multiplication. The second assumption will be a good approximation provided the current does not vary by a large percentage in the time $\tau_1 = x_1/v$ required for carriers to cross the multiplication region. This will not be so at large amplitudes. However, the errors for rising and falling currents tend to cancel, so the equation is right on the average. As we shall see, at large amplitudes only the average is involved.

2.5 *Differential Equation for $I_0(t)$*

Adding (10) and (11), using the assumption $I_p + I_n = I_0(t)$ and integrating from $x = 0$ to $x = x_1$ gives

$$\tau_1 \frac{dI_0}{dt} = -[I_p - I_n]_0^{x_1} + 2I_0 \int_0^{x_1} \alpha\, dx, \tag{12}$$

where $\tau_1 = x_1/v$ is the transit time across the multiplication region. The boundary conditions are shown in Fig. 5. The hole current at $x = 0$ consists entirely of the reverse saturation current I_{ps} of holes thermally generated in the n^+-region; these have moved to the n^+p-junction by diffusion. Thus at $x = 0$, $I_p - I_n = 2I_p - I_0 = 2I_{ps} - I_0$. At $x = x_1$ the electron current consists of the reverse saturation current I_{ns} of electrons thermally generated both in the space-charge region and in the p^+-region, so $I_p - I_n = -2I_{ns} + I_0$. With these boundary conditions, (12) becomes

$$\frac{\tau_1}{2}\frac{dI_0}{dt} = I_0\left(\int_0^{x_1} \alpha\, dx - 1\right) + I_s. \tag{13}$$

In the *dc* case I_0 is the direct current I, so this reduces to (8).

The condition for breakdown is $\int_0^{x_1} \alpha\, dx = 1$. If a field that satisfies this is suddenly applied, I_0 will increase linearly at a rate of $2I_s/\tau_1$ and become infinite. If a larger field is applied, I_0 will approach infinity exponentially. For a smaller field I_0 will approach a finite value.

The integral $\int \alpha\, dx$ over the multiplication region depends very little on carrier space-charge. Hence it will be negligibly affected by the small differences in carrier distribution between the ac and dc cases. We can therefore combine (9) and (13) to obtain a differential equation

$$\frac{\tau_1}{2}\frac{d}{dt}\ln I_0 = (E_0/E_c)^{m+1} - 1 + \frac{I_s}{I_0} \tag{14}$$

relating the current $I_0 = I_0(t)$ in the multiplication region to the peak field $E_0 = E_0(t)$.

In most practical cases the current I_0 will be so large compared to I_s that the effect of I_s can be neglected. The correction due to I_s and the detailed formulas for evaluating the effect are given in Appendix D.

2.6 *Example*

At low enough amplitudes of oscillation we can expand E_0/E_c in powers of $E_0/E_c - 1$ and retain only the linear term. Then, neglecting I_s, equation (14) becomes

$$\frac{d}{dt} \ln I_0 = \frac{2(m+1)}{\tau_1} \left(\frac{E_0}{E_c} - 1 \right). \tag{15}$$

If I_0 is to be periodic, then E_0 must be periodic and the dc bias must be such that the average E_0 is E_c. Suppose we apply a periodic voltage with the proper bias so that $E_0 = E_c + E_a \sin \omega t$, where ω is the optimum frequency π/τ. Then

$$\ln \frac{I_0(t)}{I_0(0)} = \frac{2(m+1)}{\pi} \frac{\tau}{\tau_1} \frac{E_a}{E_c} \left(1 - \cos \frac{\pi t}{\tau} \right). \tag{16}$$

Fig. 6 shows the field and $\ln I_0$ as functions of time. Suppose $\tau/\tau_1 = W/x_1 = 20$ and $m = 6$. Then even if the amplitude E_a of variation of E_0 is as small as 1.3 per cent of E_c, I_0 will vary by a factor of 10 over a cycle. Thus we can have small signals in field and voltage but large signals in current. We shall call this the intermediate range of amplitudes.

From (16) the maximum value of I_0 is seen to occur in the middle of the cycle, where $t = \tau$. Thus, if I_0 varies by a large factor, the current is generated mainly in a pulse in the middle of the ac voltage cycle, as shown in Fig. 6.

Actually the space charge of the current will affect the $E_0(t)$ curve, which will not be exactly sinusoidal for a sinusoidal voltage. However, this does not affect the conclusion that a small field and voltage signal can give a large current signal and that the current I_0 approaches a pulse as the ac amplitude increases.

2.7 *Carrier Space-Charge*

We have now dealt with the multiplication region, and obtained an equation relating I_0 and E_0. It remains to consider the rest of the space-charge region, where current generation is negligible. We shall also neglect the reverse saturation current I_{ns} of electrons, which will be negligible compared to the total current $I(x, t)$. From Poisson's equation

we shall derive another relation between I_0 and E_0. This will involve also the voltage $V(t)$ and together with (14) will uniquely determine $I_0(t)$ and $E_0(t)$ for any $V(t)$.

2.8 *Physical Picture*

The physical picture is shown in Fig. 7(a). The width of the narrow multiplication region is small compared to the total width W of the space-charge region. Therefore, in treating the current, space-charge and field distributions throughout the whole space-charge region, we shall assume that the multiplication region has zero width so that all current is generated at $x = 0$. Then $I_0(t)$ is a current of holes flowing out of $x = 0$, and the only carriers in the space-charge region are the holes.

The current $I(x, t)$ at any point x and time t is

$$I(x, t) = I(0, t - x/v) = I_0(t - x/v). \tag{17}$$

Thus the entire hole and current distributions are given in terms of $I_0(t)$.

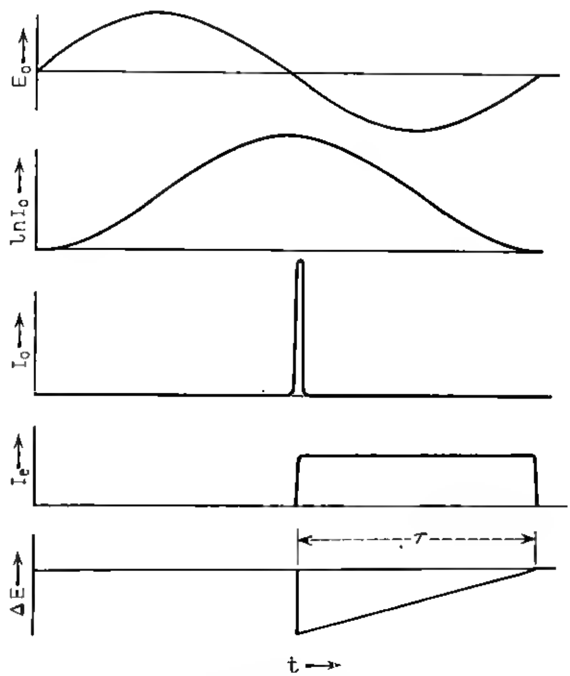

Fig. 6 — Case of a sharp current pulse.

Next consider the effect of the holes on the space charge and the field. The field distribution $E_p(x)$ at punch-through is shown by the dashed curve in Fig. 7(b). If no current flowed the field distribution for any voltage above the punch-through voltage would be simply $E_p(x)$ plus a constant determined by the voltage. The fixed negative charge in the p-material gives a drop in field across the space-charge, or depletion, layer. The positive space charge of the holes opposes that of the acceptors and hence reduces the drop in field. So, for a given voltage, E_0 will decrease as the current flowing in the space-charge region increases and flattens the field distribution.

2.9 *External Current*

The holes traversing the space-charge region give rise to a current $I_e = I_e(t)$ in the external circuit. I_e is equal to the average current flow-

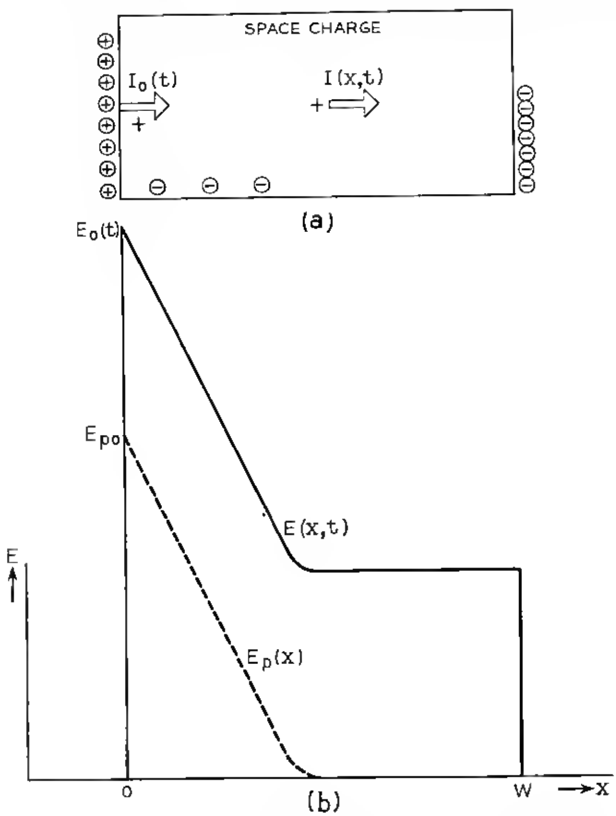

Fig. 7 — Current (a) and field (b) in the space-charge region.

ing in the space-charge region. (A rigorous and simple proof of this for the present case of plane parallel geometry is given in Appendix A.) Thus

$$\begin{aligned} I_e(t) &= \frac{1}{W}\int_0^W I(x, t)\, dx \\ &= \frac{1}{W}\int_0^W I_0(t - x/v)\, dx \\ &= \frac{1}{\tau}\int_{t-\tau}^{t} I_0(t')\, dt'. \end{aligned} \tag{18}$$

In other words, the current I_e in the external circuit is the total charge in the space-charge region divided by the transit time τ. Fig. 6 shows I_e for the case discussed at the end of Section III, where I_0 was a sharp pulse in the middle of the cycle. From (18) it follows that the average value of $I_e(t)$ is equal to the average of $I_0(t)$.

In addition to I_e, which arises from carriers moving through the space-charge region, there is a capacitative current

$$I_c = C\,\frac{dV}{dT} \tag{19}$$

flowing in the external circuit. When the voltage $V(t)$ across the diode is above the punch-through voltage, V_p, the capacity is a constant given by (2). As discussed earlier, I_c is the current required to charge and discharge the diode regarded as a capacitor. It furnishes the variation in charge at the edges of the space-charge region.

2.10 *Effect of Current on Field*

We now show how the space charge of the holes reduces the peak field for a given voltage. The stability of the device comes from the fact that current multiplication increases as E_0 increases but the current carriers give a space charge that reduces E_0.

If there were no current flowing, any increase in V above V_p would simply raise the entire field distribution by an amount $(V - V_p)/W$. Fig. 8(a) is a plot of the difference $E'(x, t) = E(x, t) - E_p(x)$ at a given time. The slope of the $E'(x, t)$ curve is determined entirely by the space charge of the holes; the effect of the fixed charge is already included in $E_p(x)$. The holes give a charge density $I(x, t)/v$. Thus, Poisson's equation is

$$\frac{\partial E'}{\partial x}(x, t) = \frac{4\pi}{\kappa v} I(x, t) = \frac{4\pi}{\kappa v} I_0(t - x/v). \tag{20}$$

The excess of V over V_p is equal to the area under the curve in Fig. 8(a). This is equal to $WE'(0, t) = W[E_0(t) - E_{p0}]$ plus the sum of the areas of a number of horizontal strips like the one shown. The area of such a strip is $(W - x)\, dE'$. So

$$V(t) - V_p = W[E_0(t) - E_{p0}] + \int_0^W (W - x) \frac{\partial E'}{\partial x}\, dx.$$

Substituting (20) into this and setting $t' = t - x/v$ gives

$$\begin{aligned} E_0(t) &= E_{p0} + \frac{V(t) - V_p}{W} + \Delta E(t), \\ \Delta E(t) &= -\frac{4\pi}{\kappa\tau} \int_{t-\tau}^{t} I_0(t')[\tau - t + t']\, dt'. \end{aligned} \tag{21}$$

Here $\Delta E(t)$ is the effect of current on the field. The quantity ΔE is always negative.

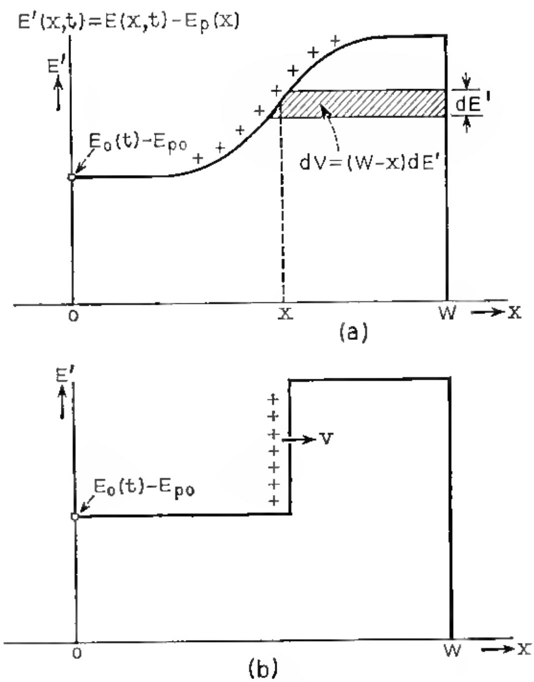

Fig. 8 — Effect of carrier space-charge on field.

2.11 *Effect of a Current Pulse*

To illustrate the physical meaning of (21) we return to the example illustrated in Fig. 6. A sharp pulse of current was generated near the middle of the cycle. The reduction $-\Delta E$ in E_0 due to this pulse is also shown in Fig. 6. If the pulse is instantaneous, $-\Delta E$ jumps at once to its maximum value and then declines to zero linearly in the time τ required for the pulse to cross the space-charge region. The physical reason for this is easily seen by reference to Fig. 8(b), which shows the same thing as Fig. 8(a) except that the carrier space-charge is concentrated at one point, that is, in a pulse. For a given voltage (area under the curve), the reduction in E_0 will decrease from its maximum value to zero as the pulse moves to the right across the space-charge region. From the effect of a single instantaneous pulse of current on E_0, the effect of any arbitrary current pulse can be found by resolving it into a series of instantaneous pulses and superposing the effects. An instantaneous pulse of current of charge δQ gives an instantaneous drop in field of $4\pi\delta Q/\kappa$.

III. ANALYSIS

We now have two equations, (14) and (21), relating the current $I_0(t)$, the field $E_0(t)$ and the voltage $V(t)$. Thus in principle the current can be found for any applied voltage. Actually the exact solution is impractical except in limiting cases. In this section we present (a) an exact solution for the linear small-signal case, (b) an approximate analysis for large amplitudes and (c) a rapidly converging iteration method that yields solutions of any desired accuracy.

3.1 *Voltage*

We shall assume that the voltage varies sinusoidally, $V(t) = V_d + V_a \sin \omega t$. This will certainly be a good assumption in the small-signal range, where the diode is linear. The cavity is linear at all amplitudes. At large amplitudes of oscillation a sinusoidal voltage gives a sinusoidal capacitative current I_c plus a conductive current I_c which approaches a square wave as the amplitude increases. Thus we are assuming that the voltage across the cavity is sinusoidal while the current contains a distribution of higher frequencies. For a cavity like that shown in Fig. 2, this assumption may be a relatively good approximation. If the cavity is tuned to the fundamental it may be almost a short circuit for the higher frequencies.

3.2 *Dimensionless Variables*

It will simplify the discussion to express everything in terms of dimensionless variables, taking as units parameters that characterize the de-

vice. For example, W would be unit length, τ, unit time, and E_c, unit field. Then the carriers would travel with unit velocity $v = W/\tau = 1$. As can be seen in Fig. 1, the voltage will be of the order of, but less than, the unit voltage WE_c. It is convenient to choose the unit charge so that $4\pi/\kappa = 1$. Then unit charge produces a unit gradient of field. Since $v = 1$, a unit charge moving in the space-charge region gives unit current. Hence unit hole current in the space-charge region produces unit field gradient, and the average current I_e in the space-charge region is equal to the total drop in field due to the carrier space-charge. The actual current will be small compared to unit current since the curent produces a drop in field that is smaller than the ac field variation, which in turn is small compared to E_c.

From the choice of units it follows that the diode has unit capacity and that optimum frequency is $\omega = \pi$. The following table summarizes the units and gives typical values for a silicon diode with $W = 10^{-3}$ cm. McKay's data,[2] plotted in Fig. 3, gives $E_c = 350$ kilovolts per cm; the effective width of the multiplication region at breakdown is taken to be 10^{-4} cm.

Quantity	Unit	Example
length	W	10^{-3} cm
time	τ	10^{-10} sec
field	E_c	3.5×10^5 volts/cm
voltage	WE_c	350 volts
current density	$\frac{\kappa E_c}{4\pi\tau} = vCE_c$	3.7×10^3 amps/cm^2
power density	$\frac{\kappa v}{4\pi} E_c^2$	1.3×10^6 watts/cm^2

The unit of power is seen to be relatively independent of W, since unit voltage goes as W and unit current as $1/W$. (A given current causes a greater drop in field across a wider space-charge region.)

3.3 *Governing Equations*

Since the peak field $E_0(t)$ varies around E_c it is convenient to define a dimensionless field

$$E(t) = \frac{E_0(t) - E_c}{E_c}. \tag{22}$$

Then, for a sinusoidal voltage variation, equation (21) becomes in dimensionless terms

$$\begin{aligned} E(t) &= E_b + V_a \sin \omega t + \Delta E, \\ \Delta E &= -\int_{t-1}^{t} I_0(t')[1 - t + t']\, dt'. \end{aligned} \tag{23}$$

Here ΔE is the effect of current on E and E_b the effect of the dc voltage V_d. Let E_{p0} be the peak field at punch-through. Then, at punch-through, E is $(E_{p0} - E_c)/E_c$. If no current flowed, any increase in dc voltage above the punch-through voltage V_p would simply raise the whole field distribution by $(V_d - V_p)/W$ (see Fig. 7). Thus the dimensionless parameter E_b is given in terms of dimensional quantities by

$$E_b = \frac{E_{p0} - E_c}{E_c} + \frac{V_d - V_p}{WE_c}. \tag{24}$$

The value of E_b will be very small compared to unity.

In equation (14) we neglect I_s and expand the right-hand side in terms of powers of $E = E(t)$. It will be sufficient to stop the expansion at E^2. Then

$$\frac{d}{dt} \ln I_0 = \frac{2(m + 1)}{\tau_1} \left(E + \frac{m}{2} E^2 \right). \tag{25}$$

The current through the diode is the capacitative current

$$I_c = C\, dV/dt = \omega V_a \cos \omega t \tag{26}$$

plus the conductive current I_e, where, from (18),

$$I_e(t) = \int_{t-1}^{t} I_0(t')\, dt'. \tag{27}$$

Average, or dc, Values

We shall let $\langle \ \rangle$ denote time averages and define I_d as the direct current $I_d = \langle I_e + I_c \rangle = \langle I_e \rangle = \langle I_0 \rangle$ where the last step follows from (27). Averaging (23) gives

$$\langle E \rangle = E_b - \frac{I_d}{2} \tag{28}$$

or $\langle \Delta E \rangle = -I_d/2$. Since $I_0(t)$ is periodic, we have, from (25)

$$\langle E \rangle + \frac{m}{2} \langle E^2 \rangle = 0. \tag{29}$$

Thus, in the small-signal limit the average field $\langle E \rangle$ vanishes, and $I_d = 2E_b$, where E_b is given by (24).

3.4 *Linear, Small-Signal Case*

When all quantities vary by small fractions of their average values, then the equations separate into a dc part and a linear ac part which is

easily solved. We now derive the impedance as a function of the direct current I_d for the optimum frequency $\omega = \pi$. The ac voltage is $V_a \sin \pi t$ or, in complex form, $V_a e^{i\pi t}$. We write $I_0 = I_d + I_{0a} e^{i\pi t}$ where I_{0a} is complex. Similar expressions can be written for I_e and E. Equations (23), (25) and (27) give the ac relations

$$
\begin{aligned}
i\pi I_{0a} &= \frac{2(m+1)}{\tau_1} I_d E_a, \\
I_{0a} &= \frac{i\pi}{2} I_{ea}, \\
i\pi(V_a - E_a) &= I_{0a} - I_{ea}.
\end{aligned}
\tag{30}
$$

From equations (30), $V_a = Z I_{ea}$ where

$$
Z = \frac{1}{2}\left[1 - \frac{\pi^2 \tau_1}{2(m+1)I_d}\right] + \frac{i}{\pi}. \tag{31}
$$

The current through the diode has two parts, I_e and I_c, where I_c is a pure capacitative current. Thus the equivalent circuit consists of a unit capacity in parallel with an element of impedance Z, where I_e goes through Z. The impedance Z consists of a fixed reactance and a resistance that varies with the dc bias. A simpler equivalent circuit emerges from considering the admittance Y of the diode. Since the capacity and the impedance Z are in parallel, $Y = i\pi + 1/Z$. Here the Q of the diode is the ratio of the imaginary part Y_i of Y to the real part Y_r. From (31)

$$
\begin{aligned}
Q &= \frac{\pi}{2}\left[1 - \frac{\pi^2 \tau_1}{2(m+1)I_d}\right], \\
Y_r &= \frac{\pi Q}{1+Q^2}, \qquad Y_i = \frac{\pi}{1+Q^{-2}}.
\end{aligned}
\tag{32}
$$

Thus Q varies linearly with τ_1/I_d and is negative for I_d less than $[\pi^2 \tau_1/2(m+1)]$. When I_d is equal to this critical value the diode becomes an open circuit for this frequency. This means that none of the alternating current generated in the multiplication region flows out of the diode. Rather it flows into the edges of the space-charge region and provides the current that charges and discharges the diode regarded as a capacitor. In other words, the unit acts as a capacitor generating its own charge internally. Hence voltage can vary with no external alternating current.

3.5 *Equivalent Circuit*

Equations (32) describe a simple equivalent circuit consisting of a fixed unit capacity in series with a conductance πQ, where Q depends on

the dc bias. Tbe equivalent circuit suggests the following practical conclusion: If the cavity is designed to act like an inductance in series with a variable load resistance, then the load resistance can be made equal to the negative resistance of the diode, so that the two resistances cancel, and the equivalent circuit consists of the fixed capacity of the diode and the inductance of the cavity. Hence at small amplitudes the resonant frequency will be independent of the dc bias.

3.6 *Sharp Pulse Approximation*

As the oscillation builds up, the behavior rapidly ceases to be linear, and it is impractical to solve the equations exactly. However, as we have seen, when the amplitude increases the current $I_0(t)$ approaches a sharp pulse and $I_e(t)$ approaches a square wave as shown in Fig. 6 for the optimum frequency $\omega = \pi$. The average current I_d is half the maximum. In the limit of a sharp pulse the problem again becomes simple. We now derive some approximate relations for this case, and show how the oscillation can be stabilized.

The power delivered P_r is

$$P_r = -\frac{V_a}{2}\int_0^2 I_e(t)\sin\pi t\,dt.$$

We can substitute (27) for $I_e(t)$ into this and reduce the double integral to a single integral by integrating by parts. The result is

$$P_r = -\frac{2V_a}{\pi}\langle I_0(t)\cos\pi t\rangle \tag{33}$$

where again the brackets denote the average over a cycle.

Thus, if the current I_0 is generated in a pulse near the middle of the cycle, where $\cos\pi t$ is negative, P_r will be negative and power will be delivered to the ac signal.

We define the Q of the diode as 2π times the ratio of the energy stored in the capacity to the energy lost in a cycle. The stored energy for unit capacity is $V_a^2/2$. The energy lost is the negative of the power delivered, P_r, times the period $2\pi/\omega = 2$. So

$$\frac{1}{Q} = \frac{4}{\pi^2}\frac{\langle I_0(t)\cos\pi t\rangle}{V_a}. \tag{34}$$

Let the pulse of current occur at a time t_1. Then in the limit of an instantaneous pulse (34) becomes

$$\frac{1}{Q} = \frac{4}{\pi^2}\frac{I_d}{V_a}\cos\pi t_1. \tag{35}$$

The phase relations will be ideal for $t_1 = 1$; that is, when the pulse occurs in the middle of the cycle. In this case $-Q$ increases with V_a for constant direct current, so the oscillation is stable.*

We now consider how t_1 depends on V_a and the dc bias. The current pulse becomes sharper as V_a increases and τ_1 decreases, as can be seen from (16). In the limit of vanishing τ_1 the pulse becomes instantaneous and the problem can be solved exactly. This is done in Appendix B. Here we give a simple physical argument that will be a good approximation so long as the duration of the pulse is small compared to a period, as will be the case in the range of practical operation.

Fig. 9 shows $E(t) = E_b + V_a \sin \pi t + \Delta E$ for the case where the current pulse occurs at an arbitrary time t_1. As illustrated in Fig. 6, a

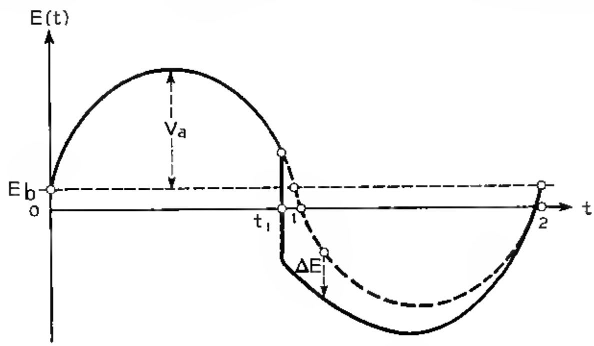

Fig. 9 — Variation of peak field with time for sharp current pulse.

current pulse causes ΔE to drop abruptly and then rise linearly to zero in a transit time. Let E_δ be the abrupt drop in E. We have already seen that the average ΔE is $-I_d/2$. Since ΔE is a triangular wave, lasting half a cycle, it follows that the maximum is four times the average; so $E_\delta = 2I_d$.

Let us consider in some detail the relatively short interval during which current is being generated. The holes can move only a short distance during this time. Hence, as seen from Fig. 8, the field will drop by an amount roughly proportional to the amount of charge generated. It will, therefore, have dropped by $E_\delta/2$ when half the charge has been generated. The pulse will be roughly symmetrical and will have reached its peak when half the charge has been generated. The current is a maximum at $E(t) = 0$ since it builds up for positive E and decreases for negative E. Thus, during the first half of the pulse E drops to zero, the

* I am indebted to J. L. Moll for pointing out the advantage of applying the bias with a constant direct current generator.

drop being $E_\delta/2$. During the second half of the pulse the field continues to drop to $-E_\delta/2$. This is shown in Fig. 9. Before the pulse occurs the field is $E(t) = E_b + \sin \pi t$. Hence in the limit of an instantaneous pulse $E_b + V_a \sin \pi t_1 = E_\delta/2$. Since $E_\delta = 2I_d$

$$I_d = E_b + V_a \sin \pi t_1 . \tag{36}$$

Eliminating $\langle E \rangle$ from equations (28) and (29) gives another relation between I_d and E_b :

$$I_d = 2E_b + \left(\frac{m}{2}\right) \langle E^2 \rangle .$$

Since E_b , V_a and $-\Delta E \leqq 2I_d$ are all small compared to unity, the only contribution to $\langle E^2 \rangle$ that cannot be neglected in comparison with I_d and E_b is $\langle V_a^2 \sin^2 \pi t \rangle = V_a^2/2$. So we have

$$I_d = 2E_b + \left(\frac{m}{2}\right) V_a^2. \tag{37}$$

Eliminating E_b between (36) and (37) gives

$$\sin \pi t_1 = \frac{I_d}{2V_a} + \frac{m}{4} V_a . \tag{38}$$

When the right-hand side is larger than unity the current cannot be an instantaneous pulse. What happens then is that the current varies almost sinusoidally and produces a space charge which keeps E small at all times. In this case the phase relation between the voltage and current makes Q positive.

In practice, we begin with a small enough bias current so that $I_d/2V_a$ has become small before the oscillation has built up to the range where the sharp-pulse approximation is valid. The effective Q of the cavity will be chosen so that the oscillation will be stabilized before $(m/4)\ V_a$ becomes comparable to unity. For example, suppose we take $I_d = V_a/2$ as discussed in Section I. Then $-\cos \pi t_1$ will have dropped only to 0.85, even for as large an amplitude as $V_a = 0.2$. For $V_a = 0.1$, $\cos \pi t_1$ would be -0.92.

3.7 *Constant dc Voltage Bias*

If a constant dc voltage bias V_d is applied then E_b is constant. As the oscillation builds up I_d increases in accord with (37).

In terms of E_b and V_a , equations (35) and (38) become

$$\frac{1}{Q} = \frac{8}{\pi^2} \left(\frac{E_b}{V_a} + \frac{m}{4} V_a\right) \cos \pi t_1 , \tag{39}$$

and

$$\sin \pi t_1 = \frac{E_b}{V_a} + \frac{m}{2} V_a . \tag{40}$$

The solid curves in Fig. 10 are plots of $-1/Q$ vs V_a for two values of E_b. In the range shown, $\cos \pi t_1$ is approximately -1. The curves, therefore, have minima at $V_a = 2\sqrt{E_b/m}$. As V_a increases, $-1/Q$ will eventually reach a maximum and begin to decrease because the term $(m/2)V_a$ in (40) will become important, and the pulse will occur too soon in the cycle. This range lies beyond that shown in Fig. 10.

At the amplitudes shown in Fig. 10, and especially in the stable range, the main error in the above approximations comes from the fact that the pulse is not sharp. We now turn to a method of obtaining more accurate results at low amplitudes.

3.8 *Iteration Method*

Equation (25) gives $I_0(t)$ when $E(t)$ is known and (23) gives $E(t)$ if $I_0(t)$ is known. Thus we can guess at $E(t)$, find the corresponding $I_0(t)$ from (25) and use it to determine a new $E(t)$ from (23), and so on.

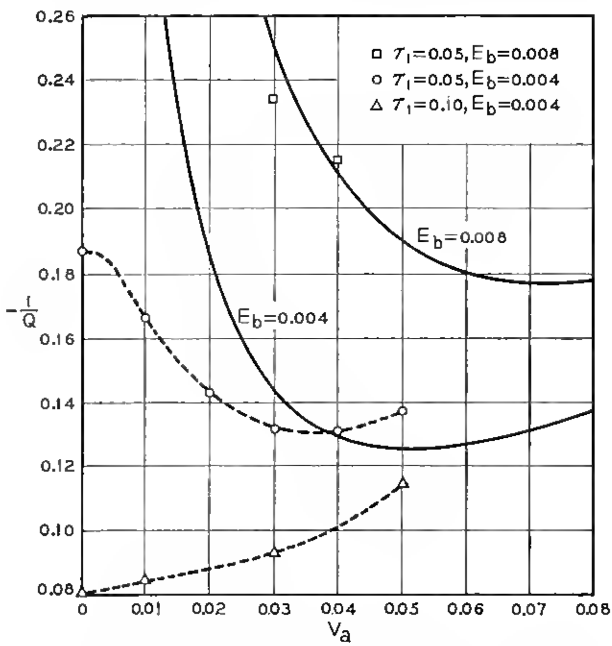

Fig. 10 — Variation of Q with amplitude for several dc voltage biases.

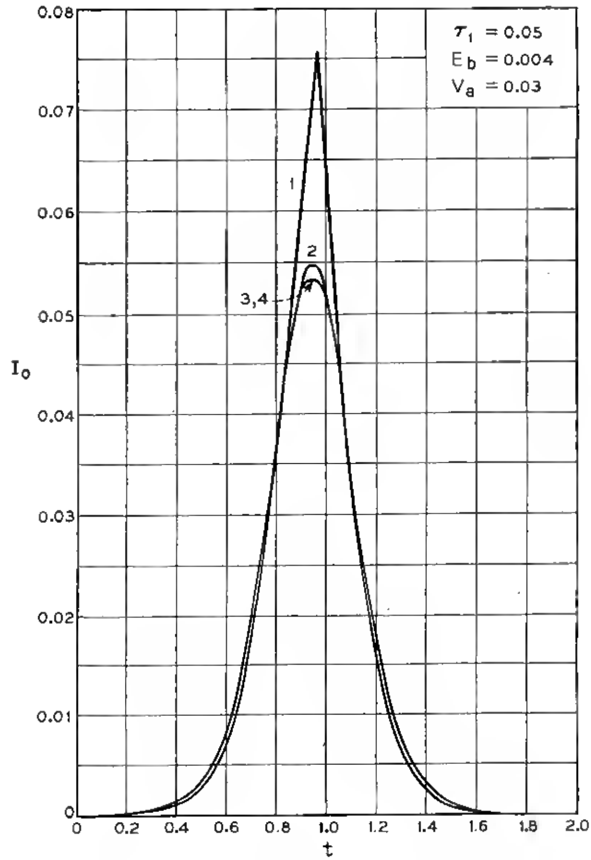

Fig. 11 — Converging solutions for $I_0(t)$.

For the initial $E(t)$ we take the instantaneous pulse solution. The procedure converges rapidly and results in solutions of any desired accuracy. In finding $I_0(t)$ from $E(t)$ in (25) a constant of integration, $I_0(0)$, is involved. What is found directly from $E(t)$ is $I_0(t)/I_0(0)$; $I_0(0)$ is so chosen that the next $E(t)$ will satisfy the condition (29) that the following $I_0(t)$ be periodic. The procedure is discussed in more detail in Appendix C.

Fig. 11 shows plots of $I_0(t)$ for $\tau_1 = 0.05$, $E_b = 0.004$ and $V_a = 0.03$. The various iterations are numbered. The values of $-1/Q$ in successive iterations were (to four places): 0.1416, 0.1330, 0.1321 and 0.1321. Fig. 12 shows $-\Delta E$ vs t for $\tau_1 = 0.05$, $E_b = 0.004$ and various values of

V_a. The corresponding $I_0(t)$ curves are shown in Fig. 13, and $I_t(t)$ curves are shown in Fig. 14.

Fig. 10 is a plot of $-1/Q$ vs V_a for various values of E_b and τ_1. The points were obtained from the iteration procedure and the solid curves from the sharp-pulse approximation. As expected, this approximation improves as V_a, E_b and $1/\tau_1$ increase. With $E_b = 0.004$ there is a minimum for $\tau_1 = 0.05$ but none for $\tau_1 = 0.1$. In the former case stable oscillations would be possible at amplitudes below about 0.035.

3.9 *Intermediate Amplitudes*

We now have solutions for the linear small-signal range and for the large amplitude range, where the sharp-pulse approximation is good. However, there remains an intermediate range where the current variation is too large for the small-signal analysis to apply but not large enough

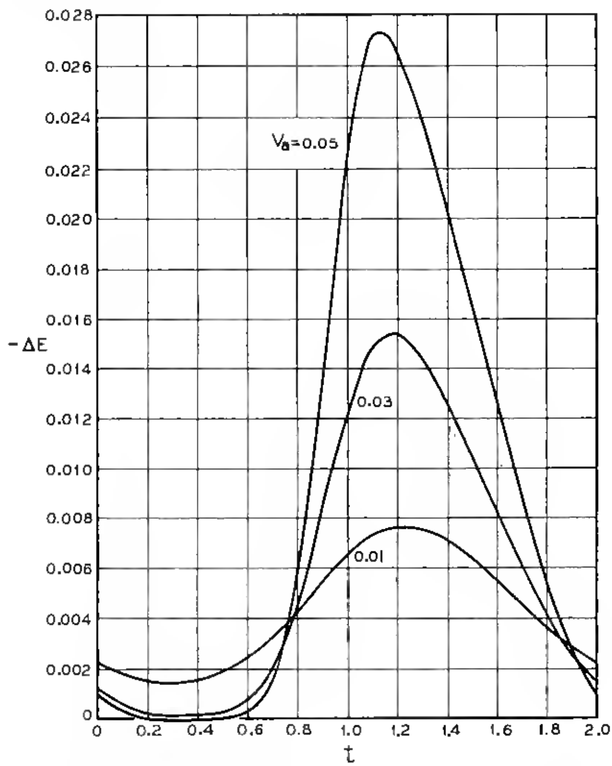

Fig. 12 — Variation of $-\Delta E$ with time for several amplitudes.

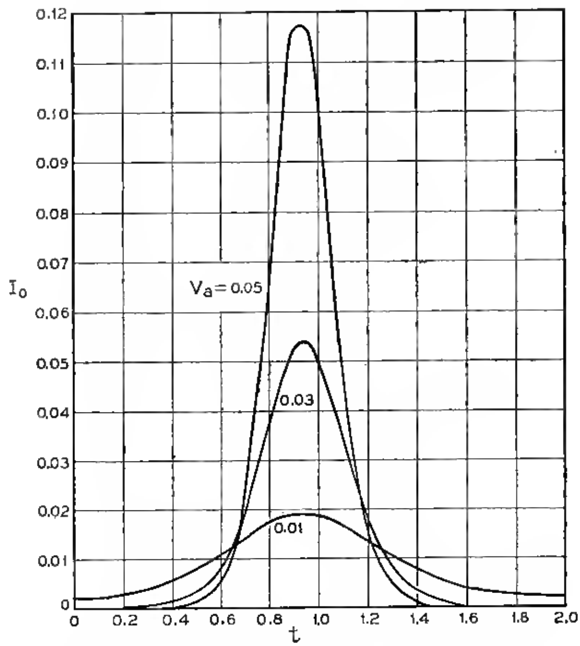

Fig. 13 — Current generated approaches a sharp pulse as amplitude increases.

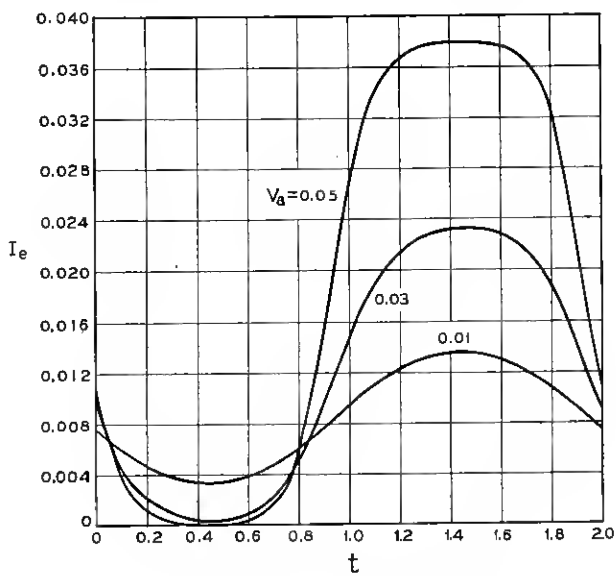

Fig. 14 — Conductive current through diode.

for the sharp-pulse approximation to be valid or even practical as a start for the iteration procedure. This range can be dealt with by an approximation that becomes more accurate as $-Q$ increases and can be within 10 per cent even for $-Q$ as low as 7. The results show how $I_0(t)$ changes from a cosine wave to an increasingly sharp pulse as V_a increases.

If Q^{-2} is small compared to unity then, in the small-signal case, $-Q = \pi V_a/I_{ea}$, where I_{ea} is the amplitude of variation of I_e. The same is true at large amplitudes if $I_e(t)$ is analyzed into a Fourier series and I_{ea} taken as the amplitude of the fundamental. In the linear small-signal range, I_{ea} is proportional to V_a so $-Q$ is constant. If I_d is kept constant as the amplitude builds up, then I_{ea} increases less rapidly than V_a. This can be seen from the fact that V_a can increase without limit while $I_e(t)$ approaches a square wave of amplitude I_d, for which $I_{ea} = (4/\pi)I_d$. Thus Q approaches $-(\pi^2/4)(V_a/I_d)$ as given by (35), for $t_1 = 1$. The phase shift, πt_1, between voltage and I_e is $\cot^{-1} Q$. We have seen that I_e is always 180° out of phase with the peak field. Therefore, if $-Q$ is large enough so that I_e and voltage are approximately 180° out of phase, the peak field is in phase with the voltage and we can use equation (16) with $E_a = V_a$. The current $I_0(t)$ is then

$$\begin{aligned} I_0(t) &= I_0(0)e^{-x(1-\cos\pi t)}, \\ x &= \frac{2(m+1)V_a}{\pi\tau_1}. \end{aligned} \tag{41}$$

This is seen to approach a pulse of increasing sharpness as V_a increases, and to reduce to the small-signal results as V_a approaches zero.

The Q can be found from (35) and (41). From (35) $1/Q = (4/\pi^2)(I_d/V_a)f(x)$, where

$$f(x) = -\frac{\langle I_0(t)\cos\pi t\rangle}{\langle I_0(t)\rangle} = -\frac{d}{dx}\int_0^\pi e^{-x\cos\theta}\,d\theta. \tag{42}$$

The function $f(x)$ is the ratio of the first to the zero order Bessels functions of pure imaginary argument. As x decreases, $f(x)$ approaches $x/2$, which means that Q is constant. As x increases, $f(x)$ bends over and approaches unity asymptotically. Values are $f(1) = 0.45$, $f(2) = 0.70$, $f(4) = 0.87$, $f(8) = 0.94$ and $f(16) = 0.97$. When I_d is constant, $Q_0/Q = 2f(x)/x$, where Q_0 is the small-signal Q. Thus the curve of $-1/Q$ vs V_a starts out flat and then decreases and approaches the form $1/V_a$ as V_a increases. This is illustrated by curves 1 and 2 in Fig. 15.

If the $-1/Q$ vs V_a curves for constant I_d are calculated from the sharp-pulse approximation [equations (35) and (38)], they have the form of the solid curves 3 and 4 in Fig. 15. As V_a decreases, the calculated $-1/Q$

goes through a maximum. For I_d below about $4/m$, the maximum would occur at about $V_a = I_d/\sqrt{2}$. However, as V_a decreases, the sharp-pulse approximation will break down. If I_d is small enough so that the small signal Q is negative, then the sharp-pulse approximation must break down before the maximum is reached, since, as we have seen, $-1/Q$ increases monotonically as V_a decreases. So the curves have the form of curves 1 and 2 in Fig. 15. However, if the small-signal Q is negative, then the curves have the form of curves 3 and 4 and the sharp-pulse approximation breaks down in the range of positive Q (shown dotted in Fig. 15). In this range the current varies roughly sinusoidally and produces a space charge that keeps the field variation small.

IV. OPERATION

In this section we consider in more detail some of the practical questions about the design and operation of the diode. In particular, we discuss the stability for both constant current and constant voltage bias, the limitations on both dc bias and ac amplitude, the effects of heating and finally the frequency dependence of the effective admittance.

4.1 *Stability at Constant Direct Current*

Fig. 15 shows the form of the variation of $-1/Q$ with voltage amplitude V_a when the bias is applied with a constant direct current generator. The oscillation will be stable at any point where $-1/Q$ is decreasing with V_a and the Q's of cavity and diode are equal in magnitude. The horizontal dashed line in Fig. 15 represents the effective $1/Q$ of the cavity. Thus the oscillation can be stabilized at the points A, B and C. The curves are numbered in order of increasing direct current I_d. If the direct current is turned on slowly compared to the response time Q_T of the diode, the oscillation will begin when I_d is slightly above the value for curve 1. Thereafter the oscillation will build up and V_a will increase as I_d increases. By slowly varying I_d we can establish a stable oscillation at any desired amplitude.

By raising I_d rapidly enough we could get onto curves 3 or 4 while the amplitude was so low as to be in the unstable range. The oscillation would then start but quickly die out.

Because of the nonlinearity of the diode, the dc voltage will vary as V_a increases at constant I_d. Eliminating E_b between (24) and (37) gives

$$I_d = 2(V_d - V_p + E_{p0} - 1) + \frac{m}{2} V_a^2, \tag{43}$$

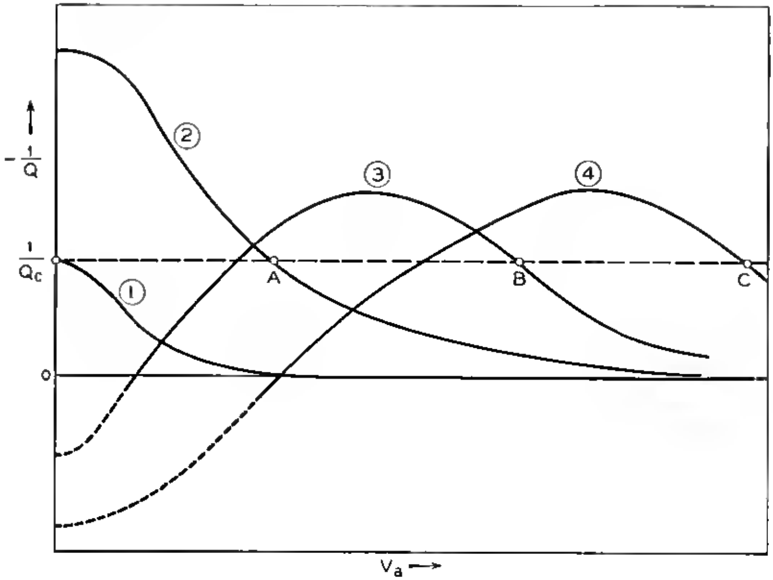

Fig. 15 — Variation of Q with amplitude (schematic) for constant direct current.

where E_{p0} is the peak field at punch through (taking E_c as unit field) and like V_p is a constant of the diode.

4.2 *Stability at Constant dc Voltage*

When the bias is applied by a constant voltage, E_b will be constant as seen from equation (24). Curves for several values of E_b and τ_1 were shown in Fig. 10. The small-signal Q is given by (32) with $I_d = 2E_b$. As E_b is raised, the small-signal Q will become negative and $-1/Q$ will rise. If Q_c, the effective Q of the cavity, is small enough the oscillation will initially be stable. For example, in Fig. 10, if $\tau_1 = 0.05$ and $1/Q_c = 0.16$, the oscillation will begin when E_b is raised to about 0.004. As E_b is further increased, the amplitude V_a will also increase. However, as seen from the figure, the oscillation will not remain stable as E_b increases unless $1/Q_c$ also increases. For example, at $E_b = 0.008$ the minimum $-1/Q$ is almost 0.18. We have seen that at large enough amplitudes $-1/Q$ reaches a maximum and decreases to zero. Thus, if E_b is increased with constant Q_c, the amplitude will suddenly jump from the range shown in Fig. 10 to a much larger value. In the range where the sharp-pulse approximation is good, $-1/Q$ is given by (37) and (40). For example, when V_a^2 is large com-

pared to $(4/m)E_b$, we have, approximately,

$$\frac{1}{Q} = \frac{4m}{\pi^2} V_a \sin 2\pi\tau_1,$$
$$\sin \pi t_1 = \frac{m}{2} V_a. \tag{44}$$

So there is a maximum in $-1/Q$ at $V_a = \sqrt{2}/m$. The maximum will be at lower amplitudes for larger E_b.

It is seen from (39) and (40) that the oscillation will always be stable when V_a^2 is between two and four times E_b/m, provided E_b is large enough that the sharp-pulse approximation is good in that range. Thus, with constant dc voltage, the oscillation can be stabilized at any amplitude, but to reach the operting point it may be necessary to vary the Q of the cavity as the bias voltage is increased.

In the remainder of this section we shall consider several effects, such as heating, which may cause the power and $-1/Q$ to decrease with increasing amplitude. These may limit the maximum power output (especially at low frequencies) and may also be used to stabilize the oscillation in the case of constant dc voltage.

4.3 *Efficiency*

We define the efficiency, ϵ, as the ratio of the ac power, P_r, to the power P_d delivered by the dc voltage or current source; therefore,

$$P_d = V_d I_d. \tag{45}$$

The power P_h that goes into heat is the difference $P_h = P_d - P_r$. Therefore,

$$P_d = \frac{P_r}{\epsilon} = \frac{P_h}{1 - \epsilon}. \tag{46}$$

Later in this section we consider the temperature rise caused by P_h.

At the optimum frequency, $\omega = \pi$,

$$\epsilon = \frac{P_r}{P_d} = -\frac{2}{\pi}\frac{V_a}{V_d}\frac{\langle I_0(t) \cos \pi t\rangle}{I_d}. \tag{47}$$

At small and intermediate amplitudes this can be evaluated by (41) and (42), which hold when $-Q$ is well above unity. In the sharp-pulse approximation

$$\epsilon = -\frac{2}{\pi}\frac{V_a}{V_d} \cos \pi t_1. \tag{48}$$

Thus, for $V_a = V_d/2$ and $t_1 = 1$, the efficiency would be over 30 per cent.

4.4 *Limitations on Bias and Amplitude*

For a desired operating amplitude we want V_d to be as small as possible to maximize the efficiency. However, as discussed in Section I, the minimum V_d is limited by the requirement that tbe field in the intrinsic region must not become negative during the negative half of the voltage cycle, and the maximum V_a is limited by the necessity of keeping the multiplication localized. We now consider these requirements in more detail.

In the analysis, we have assumed that the field in the intrinsic region is always high enough so that the carrier velocity remains constant, independent of field. However this will not be so at large enough amplitudes. The field may, in fact, momentarily become negative. This would reduce the negative resistance aud eventually destroy it. The effect of field on velocity gives both an upper limit on the allowable amplitude for a given bias and a method of stabilizing the oscillation at any desired amplitude.

The minimum field will occur in the intrinsic region and at the trailing edge of the pulse of holes advancing to the right, as illustrated in Fig. 8. The minimum field will fall below the constant-velocity range only at high enough amplitudes so that we can use the instantaneous pulse approximation. The holes will then be closely bunched as in Fig. 8(b). The pulse causes a drop in field equal to the current I_e, which is $2I_d$ during the half cycle that the pulse is moving. The field E_i immediately behind the pulse is, therefore,

$$E_i(t) = V_d - V_p + V_a \sin \pi t - 2I_d(1 - t + t_1). \tag{49}$$

This has a minimum at a time t_2 where

$$\cos \pi t_2 = -\frac{2}{\pi}\frac{I_d}{V_a}, \qquad \pi < \omega t_2 < 3\pi/2. \tag{50}$$

Equation (38) gives t_1, and (43) gives I_d in terms of $V_d - V_p$ and V_a. Thus for a given diode the minimum $E_i = E_i(t_2)$ is determined by the voltage bias and amplitude.

In general, the average field will be well above the range where velocity depends on field. Hence we can take $E_i(t_2) = 0$ as the critical condition where the drop in velocity in the negative half of the cycle begins to reduce the power appreciably. As V_a increases, the minimum E_i decreases. If the direct current is held constant, so that the oscillation is stable, then we will want $V_d - V_p$ to be large enough to build up V_a to the desired amplitude without reducing E_i below zero and thereby

losing power. However, V_d should be made as low as possible to maximize the efficiency. Thus $E_i(t_2) = 0$ is an optimum condition for operation at constant direct current. On the other hand, if we wish to operate with constant dc voltage and in the range which would be unstable for constant velocity, then V_d should be low enough so that the velocity variation will come in at the desired amplitude and stabilize the oscillation; so the condition $E_i(t_2) = 0$ would not only be optimum but necessary.

We have seen that $E_i(t_2)$ depeuds on $V_d - V_p$, V_a and the peak field E_{p0} at punch-through. Thus for a given diode $E_i(t_2) = 0$ gives a relation between $V_d - V_p$ and V_a. The quantity V_a and the corresponding $V_d - V_p$ could then be chosen to maximize the power. However, an upper limit on the voltages is set by the necessity of localizing the multiplication. The field throughout the intrinsic region must be well below the breakdown field, especially when the current is flowing. From Fig. 8(b) the field at the leading edge of the current pulse is $V_d - V_p + V_a \sin \pi t + 2I_d(t_1 - t)$. This determines how many hole-electron pairs the pulse of holes will produce in moving across the space-charge region. The holes produced join the pulse and add to the power. However, the electrons moving in the opposite direction will disturb the phase relations.

In the arguments above V_d occurs only in the combination $V_d - V_p$. It is therefore desirable to make V_p as small as possible so that V_d is small and ϵ large.

Both the voltage V_p and the peak field E_{p0} at punch-through are determined by the impurity distribution in the p-region. The amplitude at which the diode is designed to oscillate will fix the choice of E_{p0}. The larger the desired amplitude, the smaller E_{p0}. However, if E_{p0} is too small the multiplication will not be localized.

4.5 *Example*

We may illustrate the above discussion by applying it to an actual design. In the discussion at the end of Section I we took $V_d = 2V_a = 4I_d$ as a reasonable operating condition. This would give an efficiency of about 30 per cent. To operate at an amplitude of $V_a = 0.2$ would then require that $V_d - V_p$ be no less than 0.275. For $W = 10^{-3}$ and a unit voltage of 350 volts this would mean $V_a = 70$ volts, $V_d = 140$ volts and, for optimum operation, $V_p = 44$ volts. These parameters are reasonable. In fact, both V_d and V_p could probably be slightly lower so that the efficiency would be higher.

4.6 *Heating*

The power P_h that goes into heat will cause the temperature of the diode to rise above that of the surroundings. We shall assume that the surface of the cavity is kept at constant temperature. If the radius R of the diode is small compared to the radius of the center post of the cavity, as in Fig. 2, then heat can flow away from the diode in almost all directions. The diode can be made thin enough and mounted sufficiently close to the metal so that the temperature drop in the silicon is small compared to that in the metal. Let ΔT be the difference between the diode temperature and the temperature of the surface of the cavity. Then ΔT is related to P_h and R by the formula for spreading resistance

$$\Delta T = \frac{P_h R}{4K}, \tag{51}$$

where K is the thermal conductivity of the metal. For copper, $4K = 16.7$ watts per cm per °C.

The temperature will rise to its equilibrium value in a time of about R^2/D, where D is the coefficient of thermal diffusion, which is about unity for copper.

We may now apply these results to the examples discussed at the end of Section I assuming an efficiency of $\frac{1}{3}$ so that $P_h = 2P_r$. The 50 watts of power output at 5,000 megacycles and $R = 0.03$ cm would produce a temperature rise of about 60°C. At 500 megacycles and $R = 0.3$, however, the 5 kw maximum power would raise the temperature by about 600°C. Thus at low frequencies the maximum power output in continuous operation would be limited by heating rather than by how small the impedance ωL of the cavity can be made. However, the time constant of the temperature rise for $R = 0.3$ cm would be almost a tenth of a second, so the temperature rise in pulse operation would not be serious.

4.7 *Effect of Temperature on Critical Field*

McKay[2] has found that the critical field increases with temperature. For critical fields between 250 and 500 kilovolts per cm in silicon, a change in temperature changes the critical field by 0.05 per cent per °C. Increasing the critical field effectively decreases the dc voltage and current. Thus when the diode begins to get hot it effectively reduces its bias. The heating will therefore stabilize the oscillation.

4.8 *Effect of Reverse Saturation Current*

In silicon the reverse saturation current will become important only at large amplitudes. In the small-signal range it is easily shown that I_s can be neglected if it is small compared to $\pi[\pi\tau_1/2(m + 1)]^2$ which will be of the order of 10^{-3}, or a few amperes per cm^2 for $W = 10^{-3}$ cm.

Even at large amplitudes I_s will be small compared to the average current and will have a negligible effect on the space charge. However, as V_a increases, the ratio of the maximum and minimum values of I_0 (which varies exponentially with V_a) will increase much faster than the maximum I_0. Thus the minimum I_0 becomes very small at high amplitudes. It cannot, however, fall below I_s.

If the effect of I_s is included, the equation (25) for current generation becomes

$$\frac{\tau_1}{2}\frac{d}{dt}\ln I_0 = (m + 1)\left(E + \frac{m}{2}E^2\right) + \frac{I_s}{I_0}. \tag{52}$$

As V_a increases and the minimum I_0 decreases, the term I_s/I_0 becomes important near the current minimum, and prevents the current from becoming too low. Thus I_s will be important in the equation only near the current minimum. However, by increasing the minimum I_0, I_s also increases subsequent values of I_0 and hence increases the associated carrier space-charge. The carrier space-charge will therefore shut the current off (by reducing the field) earlier in the cycle. This will reduce the delay between voltage and current and so reduce the power.

The amplitude at which I_s becomes important can be roughly estimated as follows: From (41) the ratio of maximum and minimum values of I_0 is related to V_a by

$$V_a = \frac{\pi\tau_1}{4(m + 1)}\ln\frac{I_{\max}}{I_{\min}}.$$

A formula for $I_{\max}$ is derived from the instantaneous-pulse approximation in Appendix B [equation (63)] 11. In practical cases, where I_d will be of the order of 10^{-1}, $I_{\max}$ is seen to be of the order of unity. At room temperature the reverse saturation current in silicon can be made less than a microampere per cm^2, which is a dimensionless current of around 10^{-10} in the kmc range. Thus the minimum I_0 will be greater than I_s if V_a is no larger than about 2.6 τ_1. To operate at an amplitude $V_a = 0.2$ we therefore want τ_1 to be no smaller than 0.07.

The effect of I_s is treated in detail in Appendix D and more general relations are found to replace (37) and the equations derived from it.

4.9 *Frequency*

So far we have considered only the optimum frequency $\omega = \pi/\tau = \pi$. The small-signal analysis is easily carried out for arbitrary frequency. The admittance Y of the diode is found to be

$$
\begin{aligned}
Y(\omega) &= \frac{4\omega y[(1 - \cos\omega) + i(2y - \sin\omega)]}{2y^2 - 2y\sin\omega + 2(1 - \cos\omega)}, \\
y &= \frac{\omega}{2}\left[1 - \frac{\omega^2\tau_1^{\,2}}{2(m + 1)I_d}\right].
\end{aligned}
\tag{53}
$$

This reduces to (32) for $\omega = \pi$. When $-y$ is well above unity, it can be held constant by varying I_d in proportion to ω^3. Then $-1/Q$ will vary only with the phase factor $1 - \cos\omega$. Thus by varying I_d the device can be tuned mechanically over a frequency range extending from $\frac{1}{2}$ to $\frac{3}{2}$ of the optimum frequency.

At large amplitudes the current no longer varies sinusoidally. However, as mentioned in Section III, the cavity may be almost a short circuit for frequencies higher than the fundamental. Hence, to evaluate the diode as an element in the oscillator, we can analyze the current $I_e + I_c$ into a Fourier series and retain only the fundamental. The relation of this to the ac voltage then defines the conductance, G, and capacity, C, of the diode. The results are

$$
\begin{aligned}
G &= \frac{2}{\omega V_a}\,[\langle I_0(t)\sin\omega t\rangle \sin\omega + \langle I_0(t)\cos\omega t\rangle(1 - \cos\omega)], \\
C &= 1 + \frac{2}{\omega^2 V_a}\,[\langle I_0(t)\sin\omega t\rangle(\cos\omega - 1) + \langle I_0(t)\cos\omega t\rangle \sin\omega]
\end{aligned}
\tag{54}
$$

where again the brackets denote time averages. When current and voltage are approximately 180° out of phase, we can use (42) to evaluate the time averages in (54).

In the sharp-pulse approximation the averages become $I_d \sin\omega t_1$ and $I_d\cos\omega t_1$ respectively. Equation (37), which is independent of frequency, remains valid. Since the drop E_δ in field is $(4\pi/\omega)$ times the average $\langle -\Delta E\rangle = I_d/2$, the general form of (36) becomes

$$
\frac{\pi}{\omega} I_d = E_b + V_a \sin\omega t_1 . \tag{55}
$$

Above the small-signal range the admittance $Y = G + i\omega C$ of the diode depends on voltage amplitude as well as on frequency and direct current. For stable oscillation the admittance $Y = Y(\omega, I_d, V_a)$ of the diode is equal to $-Y_c$, where Y_c is the admittance of the cavity. For a

given setting of the plunger (Fig. 2), $Y_c = Y_c(\omega)$ depends only on the frequency. Thus the frequency and amplitude of oscillation are determined in terms of the direct current by $Y(\omega, I_d, V_a) + Y_c(\omega) = 0$. When I_d is varied both V_a and ω will vary. However, it is not certain whether the variation of frequency with I_d could be made large enough for a practical frequency modulation device. Near the optimum frequency and bias the frequency will remain relatively constant as I_d and V_a vary. Thus the frequency is primarily determined mechanically by adjusting the height of the cavity (Fig. 2) and the amplitude of oscillation is determined electrically by the bias.

4.10 *Zener Current*

In sufficiently narrow junctions, where breakdown occurs, the rate of generation is an extremely sensitive function of field. So, as in secondary emission, the generation can be highly localized. The diode could be made to operate by Zener current rather than multiplication if the p region is sufficiently narrow. However, the conditions for negative resistance would be less favorable. The current I_0 generated by field emission is a function of the peak field E_0. Hence I_0 and E_0 are in phase, so all of the delay has to come from the transit time. For ideal phase relations at large amplitudes the bias should be such that the current is generated mainly in a short burst near the voltage peak. Then, if the transit time is $\frac{3}{4}$ of a cycle, the current, I_e, will flow during the last three quarters of the cycle and power will be delivered to the ac signal. However, the Q and efficiency are considerably lower than for secondary emission. The small-signal Q can be varied by varying the bias. A small change in bias will change dI_0/dE_0 drastically (since we are on the knee of the current-voltage curve). For the frequency $0.75/\tau$, the minimum $-Q$ is 25 at small signals and over 100 at large signals. The latter might be improved somewhat by increasing the frequency in relation to $1/\tau$. The efficiency could probably not be raised above about 5 per cent without ruining the Q.

An analysis of the diode operation on Zener current is discussed in Appendix E.

V. ACKNOWLEDGEMENTS

The author wishes to thank J. L. Moll, G. Weinreich, B. Senitzky and R. W. Hamming for helpful discussion and advice. The machine computations leading to Figs. 10 through 14 were programmed by Mrs. G. J. Rowe.

APPENDIX A

Derivation of I_e

We first give a simple analytical derivation and then a physical argument. A more general proof is given by Shockley (1938).

Subtract (4) from (3) and use Poisson's equation

$$\frac{\partial E}{\partial x} = \frac{4\pi q}{\kappa}(p - n + N)$$

for $p - n$. The result is*

$$\frac{\partial}{\partial x}\left[\frac{\kappa}{4\pi}\frac{\partial E}{\partial t} + I\right] = 0.$$

In other words the quantity in brackets is a function of time only and is the same, at a given time, throughout the length of the diode — not only in the space-charge region, but also in the ends. The current in the ends, or leads, is $I_c + I_e$. Let the ends be of sufficiently low resistance so that the field can be assumed to vanish there. Then

$$\frac{\kappa}{4\pi}\frac{\partial E}{\partial t} + I = I_c + I_e.$$

Averaging over the length W of the space-charge region and using $C = \kappa/4\pi W$ gives

$$C\frac{dV}{dt} + \frac{1}{W}\int_0^W I\,dx = I_c + I_e.$$

The physical argument can be illustrated by Fig. 8(b). Let a small pulse of charge δQ be generated at $x = 0$. It causes a drop in field $\delta E = (4\pi/\kappa)\delta Q$. If the field at the edges remains constant as the pulse moves, the voltage will drop at a rate $v\delta E = (4\pi/\kappa)v\delta Q$. Therefore, if the voltage is to remain constant, a current δI_e must flow in the

* G. Weinreich (private communication) has pointed out that this is a special case of the general three-dimensional result

$$\operatorname{div}\left[\frac{\kappa}{4\pi}\frac{\partial \vec{E}}{\partial t} + \vec{I}\right] = 0$$

which follows from

$$\operatorname{curl}\ \vec{H} = \frac{\kappa}{c}\frac{\partial \vec{E}}{\partial t} + \frac{4\pi}{c}\vec{I}$$

and div curl = 0.

external circuit and increase the voltage by adding charges to the edges of the space-charge region. The rate of voltage increase will be $\delta I_e/C$. Setting this equal to $v\delta E$ gives $\delta I_e = \delta Q/\tau = v\delta Q/W$.

APPENDIX B

Instantaneous Pulse

The equations can be solved exactly for any amplitude and bias in the limiting case of $\tau_1 = 0$. As we have seen, the current I_0 approaches an instantaneous pulse. This solution will give reasonable approximations to the Q, efficiency and power for actual cases if the effective duration of the current pulse is a small fraction of a cycle. We let the frequency ω be arbitrary but less than $2\pi/\omega$, so that only one current pulse is flowing at a time. The solution is completely specified by the time t_1 at which the pulse and the discontinuity in field occur and the magnitude E_δ of the discontinuity. From $\langle \Delta E \rangle = -I_d/2$ and the form of $\Delta E(t)$,

$$E_\delta = \frac{4\pi}{\omega} \langle - \Delta E \rangle = \frac{2\pi}{\omega} I_d. \tag{56}$$

The condition that I_d be periodic is

$$\langle E \rangle + \frac{m}{2} \langle E^2 \rangle = 0 \tag{57}$$

where $E = E_b + V_a \sin \omega t - E_\delta(t - t_1)$ in the internal from t_1 to $t_1 + 1$ and $E = E_b + V_a \sin \omega t$ at all other times. Thus (57) is a quadratic equation for E_δ as a function of E_b, V_a and t_1. Only the smaller of the two roots is meaningful. When E_b and V_a are small compared to unity (57) reduces to (37) and is independent of t_1.

To obtain another relation between t_1 and E_δ we solve (23) and (25) and find the relation between I_0 and E during the pulse. In the limit of an instantaneous pulse, t in equation (23) does not vary during the pulse, so $dE = -I_0\, dt$. We can use this to eliminate dt in (25). Then integrating we have the relation

$$I_{\max} - I_0 = \frac{(m + 1)}{\tau_1} E^2 \left[1 + \frac{m}{3} E \right] \tag{58}$$

for $I_0(t)$ as a function of $E(t)$ during the pulse. The values of E at the beginning and end of the pulse are found by setting $I_0 = 0$ in (58). Let E_1 be the field at the beginning of the pulse. Then

$$E_1 = E_b + V_a \sin \omega t_1 , \tag{59}$$

and the maximum I_0 is

$$I_{\max} = \frac{(m + 1)E_1^2}{\tau_1}\left(1 + \frac{m}{3} E_1\right). \tag{60}$$

Neglecting terms in E_1^2 in comparison with unity, (58) gives

$$E_\delta = 2E_1\left(1 + \frac{mE_1}{6}\right). \tag{61}$$

Combining this with (57) and (60) we have

$$E_b + V_a \sin \omega t_1 = \frac{\pi}{\omega} I_d\left(1 - \frac{m\pi}{6\omega} I_d\right). \tag{62}$$

In practical cases, the term $(m\pi/6\omega)I_d \approx I_d$ can be neglected in comparison with unity. Then (63) becomes (55). This, together with the quadratic (57), determines E_δ and t_1 for any E_b and V_a.

From (60) and (61),

$$I_{\max} = \frac{(m + 1)E_\delta^2}{4\tau_1} = \frac{\pi^2(m + 1)}{\omega^2\tau_1} I_d^2. \tag{63}$$

Thus the effective duration of the pulse $\Delta t = I_d/I_{\max}$ is

$$\Delta t = \frac{\omega^2\tau_1}{\pi^2(m + 1)I_d}. \tag{64}$$

So the pulse becomes sharper as τ_1 decreases and I_d increases. We can estimate the accuracy of the instantaneous pulse approximation by comparing Δt with the period $2\pi/\omega$.

APPENDIX C

Iteration Method

The procedure in detail is the following: Each iteration goes from an $E(t)$ through an $I_0(t)$ to a new $E(t)$. In the first iteration we have to guess at $E(t)$. The procedure has been to begin with the $E(t)$ corresponding to the instantaneous current pulse and illustrated in Figs. 6 and 9. The magnitude E_δ of the discontinuity is chosen to satisfy

$$\langle E + (m/2)E^2\rangle = 0.$$

The time t_1 when the discontinuity occurs is found from (40). Putting this $E(t)$ into (25) and integrating gives a periodic $I_0(t)/I_0(0)$ from which, using (23), we find a function $F(t) = -\Delta E(t)/I_0(0)$. Putting $E = E_b + V_a \sin \pi t - I_0(0)F(t)$ into $\langle E + (m/2)E^2\rangle = 0$ gives a quadratic equation

for $I_0(0)$. Only the smaller of the two roots is meaningful. From the known $F(t)$ and $I_0(0)$ we have a new $E(t)$. Since it satisfies $\langle E + (m/2)E^2 \rangle = 0$, the $I_0(t)/I_0(0)$ calculated from it in the following iteration will be periodic.

To plot $I_0(t)$ in Fig. 11 we have determined both $I_0(t)/I_0(0)$ and $I_0(0)$ from the same $E(t)$. The quantity $I_0(t)/I_0(0)$ is found from (25) and $I_0(0)$ is chosen to satisfy $\langle I_0 \rangle = -2\langle \Delta E \rangle$. This relation is automatically satisfied if $\Delta E(t)$ has been determined from $I_0(t)$ using (23). However, only in the exact solution is it satisfied if $I_0(t)$ was determined from $\Delta E(t)$. To determine $I_0(0)$ from $\Delta E(t)$ we have

$$I_0(0) = \frac{-2\langle \Delta E(t) \rangle}{\left\langle \dfrac{I_0(t)}{I_0(0)} \right\rangle}$$

where $I_0(t)/I_0(0)$ is found from $\Delta E(t)$ using (25) with $E = E_b + V_a \sin \pi t + \Delta E$. The value of $I_0(0)$ found in this way, in each iteration, was compared with the value that makes the following $I_0(t)$ periodic. When the two values agreed within a specified amount, usually taken to be about one per cent, the iteration procedure was terminated. The procedure was programmed on an I.B.M. 650 Magnetic Drum Calculator. The machine would give a solution of the required accuracy in about four iterations, or forty minutes, on the average.

In one case, where the current peak occurred at $t_1 = 0.92$, the iteration procedure was repeated starting with $t_1 = 0.85$. The peak in successive iterations moved from 0.85 to 0.92. In other words, the final result was independent of the initial t_1. However, more iterations were required for a poor initial choice.

APPENDIX D

Effect of I_s

It is convenient to consider the solution for a single cycle extending between two current minima. The effect of I_s will be important only near the ends of the cycle where I_0 is small and the term I_s/I_0 cannot be neglected in (52). Let the solution be $f(t)I_0(t)$ where $I_0(t)$ is a solution for $I_s = 0$ and is the correct solution during most of the cycle. Then $f(t)$ will be unity during most of the cycle and will rise sharply at the beginning and end of the cycle. Since the current is always positive, each rise in $f(t)$ must be less than unity. The condition that the current be periodic is $\Delta \ln f + \Delta \ln I_0 = 0$, where Δ denotes the change during a cycle. The change in $\ln I_0$ is found by the same procedure that led to

(37). If Δf is the total change in f in a cycle we have, instead of (37),

$$\ln\left(\frac{2+\Delta f}{2-\Delta f}\right) = \frac{4\pi(m+1)}{\omega\tau_1}\left[E_b - \frac{I_d}{2} + \frac{m}{4}V_a^{\,2}\right]. \tag{65}$$

Replacing $I_0(t)$ by the correct solution $f(t)I_0(t)$ in (52) gives

$$\frac{df}{dt} = \frac{2}{\tau_1}\frac{I_s}{I_0}. \tag{66}$$

To determine Δf we need to solve this only in the short interval near the current minima where f is changing. In practical cases the current will have stopped flowing slightly before the end of the negative half of the cycle. Hence when f is changing, $E + (m/2)E^2$ will vary as $V_a\omega t$ and

$$\ln\left[\frac{I_0(t)}{I_{\min}}\right] = \frac{(m+1)V_a\omega t^2}{\tau_1} \tag{67}$$

where $I_{\min}$ is the minimum of $I_0(t)$ and $t = 0$ is taken at the minimum. Substituting (67) into (66) and integrating over the short interval where f is changing gives

$$\Delta f = \frac{2I_s}{I_{\min}}\sqrt{\frac{\pi}{(m+1)V_a\omega\tau_1}}. \tag{68}$$

The ratio of $I_{\min}$ to $I_{\max}$ is found by integrating (25) from the current minima at $E = 0$ to the maximum at $t = t_1$. At the amplitudes where the effect of I_s is important we can use the sharp-pulse approximation, so $E = E_b + V_a \sin \omega t$ during this interval. The quantity $I_{\max}$ is given by (63). Thus Δf is found in terms of E_b, V_a and t_1. Equation (65) now replaces (37) and together with (55) determines t_1 and I_d for a given voltage bias and amplitude.

APPENDIX E

Zener Current

The small-signal case is easily solved. The ac variations in I_0 and E_0 are proportional. At large signals it is probably a good approximation to say that (a) no current is generated until E_0 has risen to a critical value, which we shall define as E_c, and (b) thereafter I_0 will be such as to keep E_0 from rising above E_c. Thus (25) is replaced by $I_0 = 0$ for $E < 0$ and $E = 0$ for $I_0 > 0$. Equations (23) and (26) remain unchanged. The current $I_e(t)$ will flow during the following three intervals:

(A) Beginning at the time t_0 when $E_0 = E_c$, current begins to be generated and continues to be generated until a time t_1 when no more carrier space-charge is required to keep E_0 from rising above E_c, that

is, to keep E from becoming positive. The last holes to be generated will have been generated before the first have crossed the space-charge region. During this interval, I_e increases from $I_e(t_0) = 0$ to its maximum $I_e(t_1)$, while I_0 jumps at once to its maximum at t_0 and thereafter declines, reaching zero at t_1. The governing equation is

$$\tau \frac{dIe}{dt} - I_e = C \frac{dV}{dt} = CV_a \omega \cos \omega t. \tag{69}$$

This can be obtained by physical reasoning from Fig. 8 or by differentiating (23) and (26). Equation (69) is easily solved, subject to $I_e(t_0) = 0$. The time t_1 and the maximum current $I_e(t_1)$ are found from $dI_e/dt = 0$ at $t = t_1$.

(B) From the time, t_1, when I_e reaches its maximum and I_0 drops to zero, until the time $t_0 + \tau$, when the first holes to be generated reach the i-p$^+$ junction, I_e remains constant and equal to $I_e(t_1)$. A constant number of holes are moving with constant velocity across the space-charge region.

(C) Between $t_0 + \tau$ and $t_1 + \tau$ the holes generated in the first interval are flowing out at the i-p$^+$ junction. Since the first holes to be generated are the first to flow out, $I_e(t) = I_e(t_1) - I_e(t - \tau)$ during this interval.

Thus, by solving (69) for $I_e(t)$ in the interval $t_0 \leqq t \leqq t_1$ we know $I_e(t)$ throughout the cycle. Except in the first interval, $E \leqq 0$ and $I_0 = 0$.

The equations have been solved for the frequency $0.75/\tau$ or $\omega = 3\pi/2\tau$. The Q is a function of ωt_0, which is determined by the voltage bias and amplitude. The Q is negative for $13° < \omega t_0 < 90°$. The maximum $-1/Q$ is about 0.0095 and occurs at about $\omega t_0 = 40°$. The quantity $[\epsilon/(1 + \epsilon)]$ (V_d/V_a) rises almost linearly with ωt_0 from zero at 13° to 0.21 at 90°. The limitations on V_d and V_a would be about the same as for a diode operating by multiplication.

REFERENCES

1. A. G. Chynoweth and K. G. McKay, Phys. Rev., **106**, (1957), 418.
2. K. G. McKay, Phys. Rev., **94**, (1954), 877.
3. P. Wolff, Phys. Rev., **95**, (1954), 1415.
4. W. Shockley, J. Appl. Phys., **9**, (1938), 635